Non-Cryptographic Privacy Preserving Machine Learning Methods: A Review

Kevser Şahinbaş[1], Ferhat Ozgur Catak[2], Murat Kuzlu[3(✉)], Maliha Tabassum[3], and Salih Sarp[4]

[1] Istanbul Medipol University, Istanbul, Turkey
ksahinbas@medipol.edu.tr
[2] University of Stavanger, Rogaland, Norway
f.ozgur.catak@uis.no
[3] Old Dominion University, Norfolk, VA, USA
{mkuzlu,mtaba006}@odu.edu
[4] Virginia Commonwealth University, Richmond, VA, USA
sarps@vcu.edu

Abstract. In recent years, the use of Machine Learning (ML) techniques to exploit data and produce predictive models has become widespread in decision-making and problem-solving across various fields, including healthcare, energy, retail, transportation, and many more. Generally, a well-performing ML model requires large volumes of training data. However, collecting data and using it to predict behavior poses significant challenges to the privacy of individuals and organizations, such as data breaches, loss of privacy, and corresponding financial damage. Therefore, well-designed privacy-preserving ML (PPML) methods are significantly required for many emerging applications to mitigate these problems. This paper provides a comprehensive review of non-cryptographic privacy-preserving ML along with selected methods, such as differential privacy and federated learning. This paper aims to provide a roadmap for future research directions in the PPML field.

Keywords: Privacy-preserving · Machine Learning · Federated learning

1 Introduction

Privacy-Preserving Machine learning (PPML) is one of the most prominent application areas for data protection of computing operations [1]. This is especially crucial when the training sample contains sensitive or private data. Owners of such data may wish to use it to train a model but do not want to give up control over their data. PPML methods can help mitigate this risk by ensuring that the data used to train the model is not linked to personal identity. This can help protect the privacy of those predicted by the model.

This paper extensively reviews full-scale types of non-cryptographic privacy-preserving in the ML method, detailing what, where, and how privacy-preserving

A. Ortis et al. (Eds.): ICAETA 2023, CCIS 1983, pp. 410–421, 2024.
https://doi.org/10.1007/978-3-031-50920-9_32

can be provided. For this purpose, the concept of privacy-preserving ML is introduced and then reviewed in the literature, along with various methods for preserving data privacy while training ML models. The methodology is also briefly discussed to extend the literature review. Then, these methods are compared, and recommendations are provided for future work. The remaining sections of this paper are organized as follows. Section 2 presents non-cryptographic privacy-preserving machine learning methods. Section 3 provides opportunities and challenges. Section 4 examines the future research directions. Conclusions are drawn in Sect. 5.

1.1 Literature Review

This section provides the literature review on non-cryptographic privacy-preserving ML. Tables 1 and 2 summarize the current studies with their descriptions for differential privacy and federated learning methods, respectively.

Shokri et al. [2] developed a general system for learning from participants' data without disclosing private information. They created a neural network model using the "Distributed Stochastic Gradient Descent" optimization approach, training each participant independently. The method achieved success rates of 99.14% on MNIST data and 93.12% on SVHN data. To increase the security of the data and minimize the risks of leakage, a differential privacy approach has been applied by updating the parameters (adding noise). Firstly, the approach of applying differential privacy to Principal Component Analysis [3] was used for feature selection. A success rate of 73% was achieved on CIFAR-10 data and 97% on MNIST data. Chase et al. [4] developed a new method by using Secure Multi-Party Computation (SMPC) and differential privacy to protect the confidentiality of each sample in the training data used to create their neural network model.

Kotsogiannis et al. [5] offer One-Sided Differential Privacy (OSDP) that meets sensitivity masking. Their model assures that an attacker cannot considerably reduce the uncertainty about whether a record is sensitive using any technique. Bassily et al. [6] present a differential privacy type of the Stochastic Gradient Descent (SGD) method with enhanced composition and privacy amplification. For training models, Thakkar et al. [7] examine the adaptive gradient clip technique with user-level differential privacy, eliminating the requirement for comprehensive parameter tuning. Wang et al. [8] used non-IID (non-identically independently distributed) data to provide a new convergence analysis on local epoch size. In their study, a real-time control method that dynamically adjusts global aggregation frequency was developed. Yang et al. [9] present an extensive study of a secure federated learning framework in terms of definition, architecture, vertical FL, horizontal FL, and federated transfer learning. Chen et al. [10] check for inconsistencies between the global and lagged models by modifying the number of local periods to predict recession, expediting convergence, and avoiding straggler effect performance degradation. Konecny et al. [11] provide a technique of a communication-efficient FL model to decrease communication costs for methods of sketched updates and structured updates.

Table 1. Current studies in Differential Privacy

	Description
Differential Privacy	-Two levels of privacy protection -Doesn't share raw data during the learning process [12]
	-A refined analysis of privacy costs -Non-convex objectives, under a modest privacy budget [13]
	-Private Aggregation of Teacher Ensembles (PATE) is applied to preserve users' privacy -CT-Scan is affected by COVID-19 or not by comparing with CNN model [14]
	-Differentially private algorithms for convex empirical risk minimization -Optimal error rates are provided [6]
	-Applied the adaptive gradient clip method with user-level differential privacy [7]
	-One-sided differential privacy (OSDP) to protect sensitive records and for releasing count queries [5]
	-Local and Central Distinctive Privacy (LDP/CDP) techniques in FL -Decreases white-box membership inference attacks in FL [15]
	-PRECAD framework that provides both privacy and robustness for FL [16]

Hamm demonstrates the framework's performance with realistic tasks such as network intrusion detection, activity recognition, and malicious URL detection [27]. Choudhury et al. [12] illustrate the feasibility and usefulness of the federated learning framework in providing increased privacy while maintaining the global model's utility by applying 1 million patients' real-world electronic health data. Abadi et al. [13] developed a new method using stochastic gradient descent and differential privacy budget composition approaches. Noise is added to the gradient before updating the precision-limited network parameters of each training sample to preserve data. Chamikara et al. [18] propose a distributed perturbation algorithm called DISTPAB that achieves high accuracy, efficiency, attack resistance, and scalability for the privacy preserving of horizontally partitioned data. The privacy-preserving FedML demonstrates DISTPAB's perfect approach for preventing distributed machine learning privacy leaks while maintaining high data utility.

Tran et al. [28] propose a method for Privacy-Preserving ML models that can operate on a decentralized network setting without requiring a reliable third-party server and provide confidentiality of local data with low-cost communication bandwidth. They have designed a new method called a Decentralized Secure

Table 2. Current studies in Federated Learning

	Description
Federated Learning	-Find the best balance between local updating and global parameter aggregation for edge computing [17]
	-Sketched updates and structured updates [11]
	-A distributed perturbation algorithm using the asymmetry of resources of a distributed environment [18]
	-Two new summoning defense mechanisms, Krum and Trimmed Mean [19]
	-The LDP-FedSGD algorithm is used [20]
	-Privacy problems in composite learning [21]
	-Obtained small amounts of data from different sources from various hospitals and trains a global deep-learning model using blockchain-based FL [22]
	-FEDL outperforms vanilla FedAvg algorithm [23]
	-IoT data sharing, data offloading and caching, intrusion detection, localization, mobile audience detection, and IoT privacy and security [24]
	-A Distributed algorithm to develop an overall decentralized optimization framework [25]
	-Lightweight encryption protocol is performed [26]

Framework (SDTF) to protect the confidentiality of data in ML models. It aims to protect the data's privacy by supporting the parallel training process on a decentralized network without needing any third-party server. The Secure Sum Protocol is designed to safely calculate the sum of the participants' inputs in a large group. Randomization techniques and Secure Sum Protocol are combined to ensure the model-sharing process protects local models, even if two of them are confidential from honest but curious parties. This protocol aims to train a global model without leaking information about the local intermediate parameters and training inputs of the participants in the group. As a result of experiments on MNIST and UCI SMS spam datasets, the proposed method achieved a high success rate and efficiency for the created model.

Reich et al. [29] propose the method using Secure Multilateral Computing (SMC) to cover feature extraction from texts and classification with tree ensembles and logistic regression. They also make inferences about the reliability and accuracy of the solution. Ma et al. [30] present a new perspective on multilateral ML, which allows multiple neural networks to learn simultaneously and protects privacy in cloud computing, where huge volumes of training data are distributed among many parties. The authors conclude that the method meets the requirements for verifiability and confidentiality.

Kumar et al. [22] provide a model that obtains small amounts of data from different resources and uses blockchain-based federated learning for training a global ML method. Findings from the study show good performance. Liu et al. [31] suggest the use of federated learning for COVID-19 data training. They also compare the results of ResNet18, MobileNet, COVID-Net, and MoblieNet, four popular models with and without a federated learning method. Chaudhuri et al. focused on the classification problem of a deep neural network model where the training data consists of sensitive information [32]. They designed a method that aims to protect confidentiality in classifiers. The study used the approach of minimizing the average estimation error on the training data while determining the predictive value for each training sample by the classifier. They also used differential privacy methods on sensitive data to protect privacy. Other research on ML with differential privacy includes [33–35].

2 Non-Cryptographic Privacy-Preserving Methods

There are a variety of different methods that can be used to preserve the privacy of data while training ML models. These privacy-preserving methods are used to protect the privacy of individuals whose data is being used to train the model. These methods ensure that the data used to train the model is not linked to the person's identity. This section discusses widely used non-cryptographic privacy-preserving methods, which can be used to make ML algorithms more secure and protect sensitive data.

2.1 Differential Privacy

The concept of differential privacy is the core of privacy-preserving ML methods. Differential privacy (DP) was suggested by Dwork et al. [3], which establishes a sense of personal privacy and enables data analysis in ML. Then, it has become a prominent privacy protection technology. DP allows the extraction of useful information from a dataset without revealing any personally identifiable information about the individuals in the database, as illustrated in Fig. 1. DP is the foundation for ML and other encryption schemes that protect privacy. It is also an anonymization approach that can improve ML and mitigate privacy issues. DP can be used to generalize the ML process to mask the effects of specific input data and provide differential privacy concerning individuals, resulting in a verifiable guarantee of privacy [13]. A differential privacy method has been implemented to train data for ML algorithms based on the Stochastic Gradient Descent technique (SGD), an iterative process for incremental gradient updates to minimize a loss function.

DP is also applicable, especially for group SQL queries involving count, average, sum, maximum, minimum, and median. It can increase the privacy of the dataset by adding random noise to the query results.

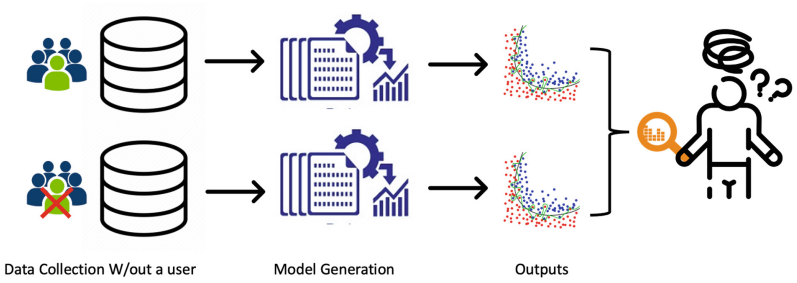

Fig. 1. The architecture of differential privacy (DP) overview

2.2 Federated Learning

Federated Learning (FedML or FL) was developed in 2016 as an efficient privacy-preserving ML technique. In this approach, many clients train their models cooperatively in a distributed environment managed by a central server while the training data is kept locally to protect privacy [9,36]. Figure 2 illustrates the general FL overview. FL enables the decentralization of ML processes by controlling the risk of compromising datasets and identity privacy as the participant limits the information exposed to datasets. Traditional centralized ML introduces system privacy issues and costs, which can be mitigated through FL. The convergence of non-IID data and communications in federated learning scenarios has been a common concern.

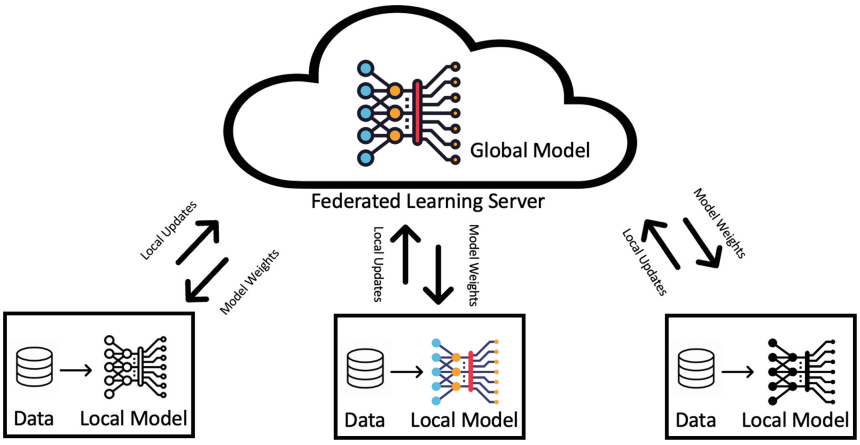

Fig. 2. General federated learning (FL) overview

FedML is branched into five types based on different aspects of federated learning as follows:

- **Federated averaging (FedAvg)**: In this method, each client trains its local model on its own data, then sends the model updates (not the data itself) to a central server. The server then aggregates the model updates from all clients and uses them to update the global model.
- **Split learning (Splitting)**: In this method, each client trains its local model on its local data and sends the output of its local model (not the data itself) to the central server. The central server then aggregates the outputs from all the clients to update the global model.
- **Federated Averaging with split learning (SplitFedAvg)**: In this method, each client sends both its local model and the output of its local model (not the data itself) to the central server. The central server then aggregates the models and outputs from all the clients to update the global model.
- **Federated Averaging with data sharing (ShareFedAvg)**: In this method, each client sends its local model and training data to the central server. The central server then aggregates the models and data from all the clients to update the global model.
- **Federated Averaging with data sharing and split learning (ShareSplitFedAvg)**: In this method, each client sends both its local data and local model to the central server. The central server then aggregates the models and the data from all the clients to update the global model.

3 Opportunities and Challenges

ML has become an integral part of many sectors, including image classification, speech recognition, natural language translation, and image analysis. These popular applications heavily rely on ML nowadays [37]. Amazon SageMaker [38], Microsoft Azure ML Studio [39], and Google Cloud ML Engine [40] are some of the known MLaaS (ML as a Service) providers. ML can be used to achieve various types of data privacy-related work. Human activity recognition (HAR) can generate massive data [41]. These datasets are from the synergy of communication [42–44] and the Medical Internet of Things (MIoT) [45,46]. These huge datasets are useful for the ML method because it enhances the study of the subject, such as the health diagnosis of patients. However, in the case of healthcare datasets, the privacy of the patient's information is sensitive. This kind of data needs to be protected from leakage. The two major algorithms for this purpose are homomorphic encryption [47,48] and differential privacy [49]. Applying these algorithms mentioned above allows a patient's data to be stored by providing privacy. Also, using Federated Learning, data sharing through ML is better and risk-free. There are remarkable outcomes in applying ML. The composition between the input and output consists of many layers. The training data consists of an individual's private information, which means the datasets can cause some risks if leaked. To prevent this, some privacy models have been adopted. Furthermore, financial companies can collect their users' information, transactions history, and other information. By applying these data to ML, it

would be easier to detect fraud. As users' data plays a vital role in enhancing datasets' accuracy, this is one reason why large companies take their users' information to train and enhance these models. Those data can be used to recognize images, label photographs to objects, etc. [50].

During the process of ML, many challenges are present. These are some of the downsides of ML models. ML is a data-driven model, which means that the more data is present, the better the results. A large amount of data is needed to feed the ML model to achieve accurate results. ML models work with the maximum amount of input data from people and try to turn it into reality by providing accurate results. Recent studies have advanced and enabled vast knowledge to be learned from. Significant achievements include efficient storage, better processing, and computing on big datasets. However, collecting a large number of datasets for a particular project can sometimes become difficult due to the unavailability of data. A low number of data might give outputs that could be more accurate, which can ruin the outcome and provide false answers.

4 Future Research Directions

Despite ML's rapid development, it still has challenges and an ever-changing room for growth. This article reviewed privacy-preserving ML methods and their latest developments. We discussed FL and different kinds of privacy-preserving mechanisms. Some potential room for growth and future research directions are discussed below.

1. To provide data privacy, several FL frameworks have been developed. However, the quality and accuracy of the data tend to degrade those adapted FL frameworks. A basic framework of FL is provided for the privacy-preserving model. A good model can be collaborated using datasets, but privacy is not guaranteed [51,52].
2. In FL, data privacy for clients lays the most important part. Gradient communication between participants and the aggregator can reveal sensitive information about the participants' datasets [53,54]. Encryption techniques such as homomorphic encryption and secret sharing can be utilized to prevent this. However, computation and communication overhead is something that encryption-based FL faces. Therefore, it is necessary to find an efficient way to stop this from occurring. Also, there needs to be a balance between the trade-off. Perturbation techniques can be utilized to protect weight and gradient updates by adding noise, but this results in degraded model accuracy and increased computational overhead. A good balance between these two conflicting performances is necessary.
3. Sensitive information can be extracted from the final model if the query results are not protected properly. Efficient solutions are needed to protect the final model. Two possible directions are: a) utilizing encryption or perturbation techniques to protect the final model against external attacks, and b) utilizing the splitting technique to personalize the model for each participant by splitting the global model [52,55,56].

4. The cost of computation and effectiveness differs in terms of privacy-preserving mechanisms. Optimization of deployment of defense mechanisms or measurements is necessary. Studies [57] show a useful guide to conduct a comprehensive investigation on diverse metrics to measure data utility and data privacy. Most studies focus on frameworks with a central server. Future research is needed to determine whether privacy attacks against an FL framework without a central server work properly or not.

5 Conclusion

A large amount of data is used in developing ML models during training and estimation, and the data used may consist of personal data. These data may include sensitive data of individuals, such as hospital and bank databases. Using this data in ML models poses security and privacy risks for data owners. Tools applied to increase the confidentiality and security of data used in ML models are given in this study. These tools are typically based-on differential privacy and federated learning for non-cryptographic privacy-preserving ML. In addition, ML-based architectures created in the literature to increase data security and privacy using privacy-preserving tools are examined, along with how and at what stage these tools are applied to the models.

References

1. Çatak, F.Ö.: Secure multi-party computation based privacy preserving extreme learning machine algorithm over vertically distributed data. In: Arik, S., Huang, T., Lai, W.K., Liu, Q. (eds.) ICONIP 2015. LNCS, vol. 9490, pp. 337–345. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-26535-3_39
2. Kubat, M.: The genetic algorithm. In: An Introduction to Machine Learning, pp. 309–329. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-63913-0_16
3. Dwork, C., Roth, A., et al.: The algorithmic foundations of differential privacy. Found. Trends Theor. Comput. Sci. **9**(3–4), 211–407 (2014)
4. Chase, M., Gilad-Bachrach, R., Laine, K., Lauter, K., Rindal, P.: Private collaborative neural network learning. Cryptology ePrint Archive (2017)
5. Kotsogiannis, I., Doudalis, S., Haney, S., Machanavajjhala, A., Mehrotra, S.: One-sided differential privacy. In: 2020 IEEE 36th International Conference on Data Engineering (ICDE), pp. 493–504. IEEE (2020)
6. Bassily, R., Smith, A., Thakurta, A.: Private empirical risk minimization: efficient algorithms and tight error bounds. In: 2014 IEEE 55th Annual Symposium on Foundations of Computer Science, pp. 464–473. IEEE (2014)
7. Thakkar, V., Gordon, K.: Privacy and policy implications for big data and health information technology for patients: a historical and legal analysis. Improving Usability, Safety and Patient Outcomes with Health Information Technology, pp. 413–417 (2019)
8. Wang, S., et al.: Adaptive federated learning in resource constrained edge computing systems. IEEE J. Select. Areas Commun. **37**(6), 1205–1221 (2019)
9. Yang, Q., Liu, Y., Chen, T., Tong, Y.: Federated machine learning: concept and applications. ACM Trans. Intell. Syst. Technol. (TIST) **10**(2), 1–19 (2019)

10. Ming Chen, Bingcheng Mao, and Tianyi Ma. Efficient and robust asynchronous federated learning with stragglers. In Submitted to International Conference on Learning Representations, 2019
11. Konečnỳ, J., McMahan, H.B., Yu, F.X., Richtárik, P., Suresh, A.T., Bacon, D.: Federated learning: Strategies for improving communication efficiency. arXiv preprint arXiv:1610.05492 (2016)
12. Choudhury, O., et al.: Differential privacy-enabled federated learning for sensitive health data. arXiv preprint arXiv:1910.02578 (2019)
13. Abadi, M., et al.: Deep learning with differential privacy. In: Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, pp. 308–318 (2016)
14. Dewang, R.K., Raven, A., Mewada, A.: A machine learning-based privacy-preserving model for COVID-19 patient using differential privacy. In: 2021 19th OITS International Conference on Information Technology (OCIT), pp. 90–95. IEEE (2021)
15. Naseri, M., Hayes, J., De Cristofaro, E.: Local and central differential privacy for robustness and privacy in federated learning. arXiv preprint arXiv:2009.03561 (2020)
16. Gu, X., Li, M., Xiong, L.: PRECAD: privacy-preserving and robust federated learning via crypto-aided differential privacy. arXiv preprint arXiv:2110.11578 (2021)
17. Wang, X. S., Huang, Y., Zhao, Y., Tang, H., Wang, X., Bu, D.: Efficient genome-wide, privacy-preserving similar patient query based on private edit distance. In: Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security, pp. 492–503 (2015)
18. Chamikara, M.A.P., Bertok, P., Khalil, I., Liu, D., Camtepe, S.: Privacy preserving distributed machine learning with federated learning. Comput. Commun. **171**, 112–125 (2021)
19. Jiang, Y., Li, Y., Zhou, Y., Zheng, X.: Sybil attacks and defense on differential privacy based federated learning. In: 2021 IEEE 20th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom), pp. 355–362. IEEE (2021)
20. Zhao, Y., et al.: Local differential privacy-based federated learning for internet of things. IEEE Internet Things J. **8**(11), 8836–8853 (2020)
21. Xu, J., Glicksberg, B.S., Su, C., Walker, P., Bian, J., Wang, F.: Federated learning for healthcare informatics. J. Healthcare Inform. Res. **5**(1), 1–19 (2020)
22. Kumar, R., et al.: Blockchain-federated-learning and deep learning models for COVID-19 detection using CT imaging. IEEE Sens. J. **21**(14), 16301–16314 (2021)
23. Dinh, C.T., et al.: Federated learning over wireless networks: convergence analysis and resource allocation. IEEE/ACM Trans. Netw. **29**(1), 398–409 (2020)
24. Nguyen, D.C., Ding, M., Pathirana, P.N., Seneviratne, A., Li, J., Poor, H.V.: Federated learning for internet of things: a comprehensive survey. IEEE Commun. Surv. Tutor. (2021)
25. Brisimi, T.S., Chen, R., Mela, T., Olshevsky, A., Paschalidis, I.C., Shi, W.: Federated learning of predictive models from federated electronic health records. Int. J. Med. Inform. **112**, 59–67 (2018)
26. Fang, H., Qian, Q.: Privacy preserving machine learning with homomorphic encryption and federated learning. Future Internet **13**(4), 94 (2021)
27. Hamm, J., Cao, Y., Belkin, M.: Learning privately from multiparty data. In: International Conference on Machine Learning, pp. 555–563. PMLR (2016)

28. Tran, A.-T., Luong, T.-D., Karnjana, J., Huynh, V.-N.: An efficient approach for privacy preserving decentralized deep learning models based on secure multi-party computation. Neurocomputing **422**, 245–262 (2021)
29. Reich, D., et al. Privacy-preserving classification of personal text messages with secure multi-party computation. In: Advances in Neural Information Processing Systems, vol. 32 (2019)
30. Ma, X., Zhang, F., Chen, X., Shen, J.: Privacy preserving multi-party computation delegation for deep learning in cloud computing. Inf. Sci. **459**, 103–116 (2018)
31. Liu, B., Yan, B., Zhou, Y., Yang, Y., Zhang, Y.: Experiments of federated learning for COVID-19 chest x-ray images. arXiv preprint arXiv:2007.05592 (2020)
32. Chaudhuri, K., Monteleoni, C., Sarwate, A.D.: Differentially private empirical risk minimization. J. Mach. Learn. Res. **12**(3) (2011)
33. Kifer, D., Smith, A., Thakurta, A.: Private convex empirical risk minimization and high-dimensional regression. In: Conference on Learning Theory, pp. 25–1. JMLR Workshop and Conference Proceedings (2012)
34. Song, S., Chaudhuri, K., Sarwate, A.D.: Stochastic gradient descent with differentially private updates. In: 2013 IEEE Global Conference on Signal and Information Processing, pp. 245–248. IEEE (2013)
35. Wu, X., Kumar, A., Chaudhuri, K., Jha, S., Naughton, J.F.: Differentially private stochastic gradient descent for in-RDBMS analytics. CoRR, abs/1606.04722 (2016)
36. Thapa, C., Arachchige, P.C.M., Camtepe, S., Sun, L.: Splitfed: when federated learning meets split learning. arXiv preprint arXiv:2004.12088 (2020)
37. Zhang, T., He, Z., Lee, R.B.: Privacy-preserving machine learning through data obfuscation. arXiv preprint arXiv:1807.01860 (2018)
38. Rauschmayr, N., et al.: Amazon sagemaker debugger: a system for real-time insights into machine learning model training. Proc. Mach. Learn. Syst. **3**, 770–782 (2021)
39. Pliuhin, V., Pan, M., Yesina, V., Sukhonos, M.: Using azure machine learning cloud technology for electric machines optimization. In: 2018 International Scientific-Practical Conference Problems of Infocommunications. Science and Technology (PIC S&T), pp. 55–58. IEEE (2018)
40. Kuzlo, I., Strielkina, A., Tetskyi, A., Uzun, D.: Selecting cloud service for healthcare applications: from hardware to cloud across machine learning. In: PhD@ ICTERI, pp. 26–34 (2018)
41. Owusu-Agyemeng, K., Qin, Z., Xiong, H., Liu, Y., Zhuang, T., Qin, Z.: MSDP: multi-scheme privacy-preserving deep learning via differential privacy. Pers. Ubiquit. Comput., 1–13 (2021)
42. Zhang, N., Peng Yang, J., Ren, D.C., Li, Yu., Shen, X.: Synergy of big data and 5G wireless networks: opportunities, approaches, and challenges. IEEE Wirel. Commun. **25**(1), 12–18 (2018)
43. Zhang, N., et al.: Software defined networking enabled wireless network virtualization: challenges and solutions. IEEE Netw. **31**(5), 42–49 (2017)
44. Chen, H., Guo, B., Zhiwen, Yu., Chen, L., Ma, X.: A generic framework for constraint-driven data selection in mobile crowd photographing. IEEE Internet Things J. **4**(1), 284–296 (2017)
45. Qin, Z., et al.: Learning-aided user identification using smartphone sensors for smart homes. IEEE Internet Things J. **6**(5), 7760–7772 (2019)
46. Qin, Z., Wang, Y., Cheng, H., Zhou, Y., Sheng, Z., Leung, V.C.: Demographic information prediction: a portrait of smartphone application users. IEEE Trans. Emerg. Topics Comput. **6**(3), 432–444 (2016)

47. Wagh, S., Gupta, D., Chandran, N.: SecureNN: efficient and private neural network training. Cryptology ePrint Archive (2018)
48. Mohassel, P., Zhang, Y.: SecureML: a system for scalable privacy-preserving machine learning. In: 2017 IEEE Symposium on Security and Privacy (SP), pp. 19–38. IEEE (2017)
49. Yin, C., Xi, J., Sun, R., Wang, J.: Location privacy protection based on differential privacy strategy for big data in industrial internet of things. IEEE Trans. Industr. Inf. **14**(8), 3628–3636 (2017)
50. Ali, S., Irfan, M.M., Bomai, A., Zhao, C.: Towards privacy-preserving deep learning: opportunities and challenges. In: 2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA), pp. 673–682. IEEE (2020)
51. Geiping, J., Bauermeister, H., Dröge, H., Moeller, M.: Inverting gradients-how easy is it to break privacy in federated learning? In: Advances in Neural Information Processing Systems, vol. 33, 16937–16947 (2020)
52. Yin, X., Zhu, Y., Jiankun, H.: A comprehensive survey of privacy-preserving federated learning: a taxonomy, review, and future directions. ACM Comput. Surv. (CSUR) **54**(6), 1–36 (2021)
53. Li, Q., Wen, Z., He, B.: Practical federated gradient boosting decision trees. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, pp. 4642–4649 (2020)
54. Hao, Y., Yang, S., Zhu, S.: Parallel restarted SGD with faster convergence and less communication: demystifying why model averaging works for deep learning. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, 5693–5700 (2019)
55. Fallah, A., Mokhtari, A., Ozdaglar, A.: Personalized federated learning with theoretical guarantees: a model-agnostic meta-learning approach. In: Advances in Neural Information Processing Systems, vol. 33, pp. 3557–3568 (2020)
56. Çatak, F.Ö., Mustacoglu, A.F.: CPP-ELM: cryptographically privacy-preserving extreme learning machine for cloud systems. Int. J. Comput. Intell. Syst. **11**, 33–44 (2018)
57. Wagner, I., Eckhoff, D.: Technical privacy metrics: a systematic survey. ACM Comput. Surv. (CSUR) **51**(3), 1–38 (2018)